Indian J.Sci.Res. 17 (1): 274-280, 2017
ISSN: 0976-2876 (Print)
ISSN: 2250-0138 (Online)

SENTIMENT ANALYSIS ON ONLINE REVIEWS USING NAÏVE BAYES CLASSIFIER METHOD AND TEXT ASSOCIATION (CASE STUDY: GARUDA INDONESIA AIRLINES PASSENGERS REVIEWS ON TRIPADVISOR SITE)

MUHAMMAD MULAJATI[a1] AND R.B. FAJRIYA HAKIM[b]

[ab]Department of Statistics in Islamic University of Indonesia

ABSTRACT

The Increasing number of population, the growing of information technology, and communication trigger the increasing of economic activity of society. Increased economic community cannot be separated from the process of ongoing transportation. Transportation is a service that is needed every day for the community. The increasing number of means of transportation has an impact on the intense competition between these types of businesses. The level of customer satisfaction is a vital aspect to survive and win business competition, one way of measuring customer satisfaction is by providing easy and convenient access for customers to submit suggestions, criticisms, opinions, and complaints. Along with the development of online media, customer reviews can be created and viewed by many people through social media. This study uses web scraping techniques to obtain Garuda Indonesia airline reviews data from the TripAdvisor site. Data obtained from the TripAdvisor site is further labeled and analyzed using the Naïve Bayes Classifier (NBC) method to classify reviews based on positive and negative sentiment categories. Furthermore, the results of sentiment classification will be analyzed by Text Mining method, the main concept is to do the widest exploration in the data of the reviews that have been obtained so that it is found an information that is considered important and can be useful for various areas of need. Sentiment classification results show more than 80% of the reviews are positive reviews with an accuracy of 82.02%. From the text association results obtained information that passengers of Garuda Indonesia majority share talk about the service, staff, seat, and food because it always appears in both positive and negative sentiment class.

KEYWORDS: Sentiment Analysis, Naïve Bayes Classifier, Text Mining, Text Association, Garuda Indonesia, TripAdvisor.

Increasing number of population, the growing of information technology, and communication trigger the increase of economic activity of society. The increase in economic society cannot be separated from the process of ongoing transportation. Transportation is a very important tool in supporting the success of development, especially in supporting the economic activities of society and regional development. With the transportation can open the way of communication between regions so that the flow of goods, services, people, and ideas as a capital for an area to move forward and develop. The development of transportation at this time can trigger the emergence of various types of public transportation for both land, sea and air transportation. With the increasing number of public transportation means impact on the intense competition between these types of business. The standard that should be emphasized in the business of transportation services, namely the service. Tjiptono (2005) states that, service (customer satisfaction) is a vital aspect to survive and win business in the competition. Based on this, then one action to satisfy the consumer is by providing services to consumers with the best. One way to measure customer satisfaction is by providing easy and convenient opportunities and access for customers to communicate suggestions, criticisms, opinions, and complaints [9].

Along with the rapid development of online media or technology, today's comments or customer complaints can be made and seen by many through social media. One of them is on the TripAdvisor site. TripAdvisor is one of the largest tourist sites in the world that help travelers optimize the potential of each trip. TripAdvisor offers advice from millions of travelers as well as a variety of travel planning options and features with a handy link to a booking tool that checks out hundreds of websites to find the best hotel, restaurant, and transportation prices. On the TripAdvisor page also provides information about the reviews of travelers about a hotel, restaurant, and some airlines. The existence of an online review on the TripAdvisor site will assist the marketing for the airline as through TripAdvisor travelers who already use some services from the airline can write down their experiences so that potential travelers who see the reviews may be interested in re-using the same services. Garuda Indonesia Airlines is one of the airlines that utilizes TripAdvisor facilities as one of its marketing facilities.

Referring to the above explanation it is necessary to review the assessment of reviews or reviews

[1]Corresponding author

provided by customers to find out how the response or response of the customer to the service (satisfaction) that has been given by the Garuda Indonesia Airlines, one of them by conducting an analysis of online review given . One method used is Text Mining (Sentiment Analysis). Sentiment analysis is an analysis that can be used to determine the nature of a comment or a review. The sentiment analysis aims to extract the attributes and components of the commented object in each document and to determine whether the comment is positive or negative [4]. With the analysis of sentiments and text mining, the management of Garuda Indonesia Airlines can know the reputation of the airline, whether viewed positive or negative. In this research the sentiment analysis process will be done by using Naïve Bayes Classifier method. Then for the extraction and exploration process the author uses descriptive statistics and associations between terms (words or topics are often discussed) are interrelated.

SENTIMENT ANALYSIS

Sentiment analysis, also called opinion mining, is a field of study that analyzes the sentiments of people's opinions, evaluations, judgments, attitudes and emotions on entities such as products, services, organizations, individuals, issues, topics, and attributes [3]. Research activities in the field of sentiments and opinion mining analysis are greatly improved due to supporting factors [4]:

1. The emergence of machine learning methods in natural language processing and information retrieval.
2. The availability of data sets for machine learning algorithms for training on the world wide web in particular, the development of review-aggregation of websites.
3. The realization of exciting intellectual challenges with the intellegence and commercial applications offered.

The term analysis of sentiments was first presented by (Nasukawa and Yi, 2003) [6], and the first term opinion mining emerged by (Dave et al., 2003) [2]. Nasukawa and Yi (2003) in his paper using natural learning processing techniques on online classifier sentiment. Dave et al. (2003) in his paper introduced the opinion mining tool, which collects opinions on a particular topic classifying them according to subjective analysis. This is done by identifying the unique nature of the problem and developing a method that automatically differentiates between positive and negative reviews.

NAÏVE BAYES CLASSIFIER

In naive bayes classification each review is represented in the attribute pair where is the first word is the second word and so on, whereas V is the set of classes [5]. At the time of classification, this method will produce the highest probability category / class (VMAP) by entering attributes <a1, a2, a3, ... an>. The VMAP formula can be seen in equation (1).

$$V_{MAP} = \text{argmax}_{vj \in V}\ P(v_j \mid a_1, a_2, a_3, \dots a_n) \quad (1)$$

Meanwhile, the bayes theorem states that,

$$P(B|A) = \frac{P(A|B) x P(B)}{P(A)} \quad (2)$$

Using this bayes theorem, then (3) can be written into

$$V_{MAP} = \text{argmax}_{vj \in V} \frac{P(a1, a2, a3, \dots an \mid v_j)}{P(a1, a2, a3, \dots)} \quad (3)$$

$P(a_1, a_2, a_3, \dots a_n \mid v_j)$ The value constant for all so that equation (1) can also be expressed as (4)

$$V_{MAP} = \text{argmax}_{vj \in V}\ P(a_1, a_2, a_3, \dots a_n \mid vj)\ x\ P(v_j) \quad (4)$$

The difficulty level of counting $P(a_1, a_2, a_3, \dots a_n \mid v_j)$ will be high because the number of term depends on the number of word position combinations multiplied by the number of classes.

The Naive bayes classifier simplifies this by assuming that within each category, each attribute is conditionally free to one another [8]. In other words

$$P(a_1, a_2, a_3, \dots a_n \mid v_j) = \prod_i P(a_i \mid v_j) \quad (5)$$

Then if equation (4) is substituted into equation (6), it will produce

$$V_{MAP} = \text{argmax}_{vj \in V}\ P(v_j)\ x \prod_i P(a_i \mid v_j) \quad (6)$$

$P(v_j)$ and the probability of the word a_i for each $P(a_i \mid v_j)$ category are calculated during the training. Where

$$P(v_j) = \frac{docs_j}{training} \quad (7)$$

$$P(a_i \mid v_j) = \frac{n_i + 1}{n + kosakata} \quad (8)$$

Where $docs_j$ is the number of documents in category j and training is the number of documents used in the training process. While n_i is the number of occurrences of the word a_i in category v_j, n is the number of vocabulary that appears in category v_j and the vocabulary is the number of unique words in all training data.

RESULT AND DISCUSSION

Descriptive Analysis

Descriptive analysis in this study aims to see a general picture of information about Garuda Indonesia airline based on visitor reviews data from the previously obtained TripAdvisor site. From the data, it can generally be described the number of incoming reviews based on the time sequence such as the figure 1.

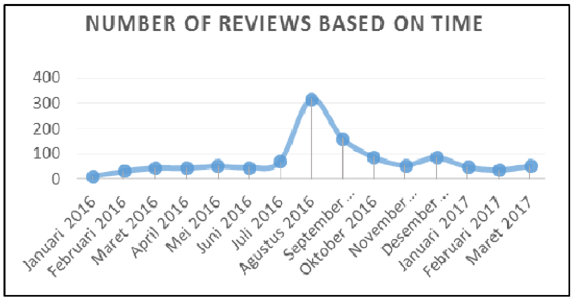

Figure 1: The number of reviews in English based on the time order

Figure 1 shows a graph of the number of visitors to the TripAdvisor site providing an english review of Garuda Indonesia airline from January 2016 to March 2017. Based on the figure it can be seen that the number of reviews in each month tends to fluctuate, the number of reviews increases significantly In August of 2016 with the number of incoming reviews of 318 reviews. The increase in the number of reviews is allegedly due to coincide with the scheduled departure of 2016 pilgrims. Garuda Indonesia is the only national airline appointed to deliver Indonesian pilgrims in 2016.

Along with the increase of passengers or flights made by Garuda Indonesia airlines, the chances of increasing the number of reviews about Garuda Indonesia airlines on the TripAdvisor site will also be greater. Then the more number of reviews that go on the TripAdvisor site, the more information that can be obtained from the review. Information in the form of passenger reviews data is of course very beneficial to Garuda Indonesia airline, because with that information the airline can indirectly know the perception and opinion of passengers on Garuda Indonesia airline, whether the perception of facilities, services, and quality so it can be used as control and materials Evaluation to a better direction.

Labeling of Sentiment Class

This stage is one of the process to get the expected corpus representation. A commonly used approach to representation of corpus is the bag-of-words model. The bag-of-words model will represent the corpus into a word and then add the same word in the corpus. In the bag-of-words representation of each word is represented by a separate variable that has numerical quantities. How to calculate the numerical quantity that is by weighting. The weighting used is dictionary-based automatic weighting (lexicon based). In this study, the lexicon dictionary used for data weighting is a dictionary composed by (Hu and Liu, 2004) which contains 6800 words. Word weighting is done by counting the frequency of word occurrences in a text document. The more often a word appears in a text document, the greater the word weight and the word is regarded as a word that strongly represents the text document [1].

In this study, the labeling process is divided into two sentiment classes, which are positive and negative sentiments by scoring. Assessment of documents entered into a class of positive or negative segmentation is determined by utilizing a collection of English words consisting of positive words is a collection of positive words and negative words is a collection of positive and negative words. The result of labeling of sentiment class is obtained by comparison of amount of data as follows:

Table 1: Comparison of the amount of data in the sentiment class

Sentiment Class	Number of Reviews
Positive	976
Negative	167
Total	**1143**

Train Data and Test Data

Train data is used by classification algorithms to form a classifier model, this model is a knowledge representation that will be used to predict new data classes that have never existed, the greater the trainer data used, the better the machine will understand the data pattern. Test data is used to measure the extent to which the classifier successfully classified correctly. The data used for training data and test data is data that already has a class label, with the amount of training data and test data has a ratio of 80%: 20%. Although extensive research has not been done in the optimal ratio selection between these data sets, there are some common practices in selecting the size of this data set [7]. Based on Paretto Principle, the commonly used ratio is 80:20 for data sets training and testing. Comparison of the amount of training data and test data can be seen in Table 2. below:

Table 2: Comparison of training data and test data

Classification	Amount	Train Data (80%)	Test Data (20%)
Positive	976	781	195
Negative	167	134	33
Total	1143	915	228

Based on Table 2., with comparative data and test data of 80%: 20%, of a total of 1143 English-language review data, 915 data were used as training data and 228 data as test data.

Classification by Naïve Bayes Classifier Method

The classification process is done by making machine learning using training data and random test data. This research uses confusion matrix method in evaluation process. Confusion matrix is one of the most important tools in evaluation method used in machine learning which usually contain two or more categories [5]. Each element of the matix shows the number of test sample data for the actual class represented in row form while the column describes the predicted class. To perform a model evaluation, this experiment was performed by iterating on the dataset as cross validation to find the best predictive accuracy value. From 100 times iteration done using Naïve Bayes Classifier method obtained the highest accuracy level that is at the 45th iteration that is equal to 84% but still have the value of recall and low precision that is equal to 6% and 29% and have confusion matrix less good. Accuracy rate calculation results are obtained from the amount of test data that is classified correctly compared with the total of all data tested. After the observation, the best classification model is the 67th iteration with 82% accuracy, 21% recall value, and 32% precision value. Table 3 below illustrates the confusion matrix results of two sentiment class predictions with the best accuracy, recall, and precision values.

Table 3. Confusion matrix results

Prediction	Actual		Class Precision
	Positive	Negative	
Positive	180	26	87,38%
Negative	15	7	31,82%
Class Recall	92,31%	21,21%	
Accuracy 82,02%			

Based on Table 3., using the Naïve Bayes Classifier method, the predicted results showed that in the positive class, from 195 positive reviews, there were 180 well-classified reviews and there was a prediction error of 15 reviews that entered into negative reviews to obtain a precision score for the positive class Amounted to 87.38%. While in the negative reviews, from the total of 33 reviews there are 7 reviews that have been correctly classified as negative reviews and there are 26 prediction errors that fall into positive reviews, resulting in negative class precision value of 31.82%. Then from the confusion matrix value obtained an accuracy of 82.02%, which means that of 228 tested review data, there are 187 reviews that are correctly classified by the model Naïve Bayes Classifier (NBC).

Visualization and Association

Visualization is done on each class of sentiment class. The purpose of visualization is to extract information in the form of topics most often discussed / reviewed by passengers Garuda Indonesia airline, so that from the many text of existing reviews, can be taken information that is considered important and sought the association between words that most often appear simultaneously, so Able to strengthen the search for such information. The following describes the visualization results and word associations of each class classification of sentiments.

Positive Reviews

Positive review data used is labeled data that is done using either lexicon dictionary or manually. The extraction of information on positive reviews is done repeatedly to get information about Garuda Indonesia airline passenger reviews which are most frequently reviewed / discussed. The positive reviews are identified by the frequency of words in the review, the following are visualizations of information extracted results obtained from visitor reviews with positive review classifications.

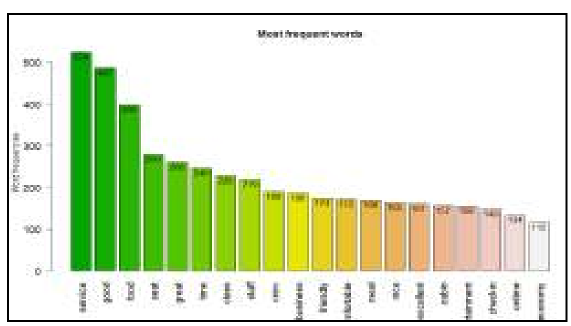

Figure 2: The word most often appears in positive class

Based on the positive review classification results, from the number of positive reviews as many as 976 reviews, obtained the most common words such as "service" with 524 times, "good" 487 times, "food" 398 times, and so on. The words that appear as in Figure 2 are words that have positive sentiments and are the topics of conversation most visited by visitors. The words are then used as a basis for finding associations with other words, so that better information can be obtained. Collection of words that often appear can also be displayed in the form of wordcloud as shown in Figure 3.

Figure 3: Wordcloud for positive reviews

Based on wordcloud visualization can be viewed more clearly the topic and positive words that visitors often use in providing reviews. The larger word size on wordcloud describes the higher the frequency of the word, meaning that the more visitors use the word as a topic of conversation or a positive rating in the review. Furthermore, the search between the associations of words that often appear simultaneously and obtained the following results:

Table 4: The word association in the positive sentiment class

service		food	
good	0,19	good	0,20
excellent	0,18	impressive	0,15
exceptional	0,14	freshly	0,14
executive	0,13	stunning	0,14
faultless	0,13	delicious	0,13
quality	0,12	classy	0,12
exclusively	0,12	flavors	0,12
		variety	0,11
		great	0,10
seat		**staff**	
comfortable	0,26	friendly	0,19
exclusively	0,25	great	0,17
length	0,25	helpful	0,14
large	0,20	assistance	0,14
simple	0,15	professional	0,13
elite	0,15	rushing	0,12
comfy	0,12	wonderful	0,12
easy	0,12		
roomy	0,12		

Based on Table 4., there are some word associations in positive class classification. The process of extraction of information with associations is done repeatedly by filtering words that are related to other words and based on word relevance to the topics covered. From Table 5.10 above, when viewed by word association relating to the word "service", can be obtained information about service or service very good, extraordinary, exclusive and perfect.

The words associated with the word "food" also provide information on diverse foods, ranging from good, fresh, tasty, varied, classy and impressive food.

The words associated with the word "seat" provide information about comfortable, exclusive, elite, long, spacious, simple and roomy seating.

The words associated with the word "staff" provide information about the performance of Garuda Indonesia airline staffs are considered friendly, professional, great, considerate, caring and hurry to help passengers.

Negative Reviews

The extraction of information on negative reviews is repeated over and over again to get information about the most frequently reviewed / discounted Garuda Indonesia airline reviews. Based on labeling results, visitors' negative reviews on airlines are quite small when compared to the number of positive reviews. From a total of 1143 reviews, only 167 negative reviews were identified. This indicates that the majority of Garuda Indonesia airline passengers have a good perception of the airlines. The results of information extraction in the form of negative reviews are identified by the frequency of words in the review, but they are also based on word relevance with topics that refer to negative sentiments. The following is a visualization of the information extraction results obtained from passenger reviews with a negative review classification.

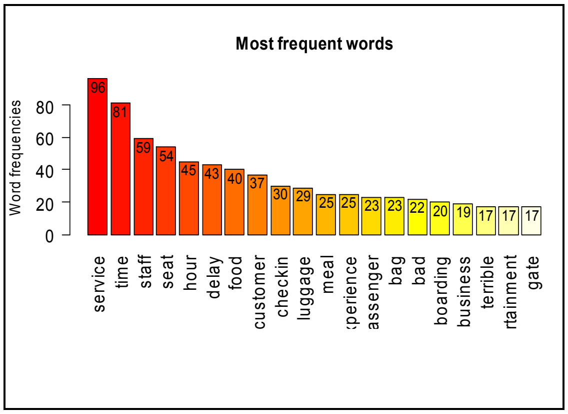

Figure 4: The word most often appears in negative class

Based on the negative review classification results, some of the most common words with relevant topics are negative sentences such as "service" with frequency 96 times, 81 time "time", "staff" as much as 59 times, "seat" 54 times, "hour" 45 times, and so on. The words that appear as in Figure 4 are words that have negative sentiments in English and are the topics most talked about by passengers. The words are then used as a basis for finding associations with other words, so that more accurate negative sentiments can be obtained. Collection of words that often appear can be displayed in the form of wordcloud as shown in Figure 5.

Figure 5: Wordcloud for negative reviews

Wordcloud visualization in Figure 5 provides a clearer picture of the topics and negative words that visitors often use in providing reviews. Some topics that are often discussed by visitors are about service, time, staff, seat, hour and so on. Furthermore, the search between the associations of words that often appear simultaneously and obtained the following results:

Table 5: The word association in the positive sentiment class

service		staff	
distance	0,46	catch	0,43
catering	0,46	missed	0,41
postponed	0,46	apologies	0,41
price	0,36	confused	0,40
full	0,35	misunderstanding	0,40
ridiculous	0,31	unfriendly	0,33
worse	0,29	arguing	0,29
anger	0,21	uninformed	0,29
convoluted	0,21	disinterested	0,29
fiasco	0,21	undelivered	0,29
pathetic	0,21	lose	0,26
confusing	0,17	mess	0,26
compensation	0,16	communication	0,19
seat		**food**	
begging	0,38	inedible	0,52
distance	0,38	pasta	0,51
separate	0,38	dried	0,44
full	0,34	hungry	0,44
cramped	0,28	local	0,44
worse	0,27	serving	0,44
smaller	0,27	western	0,29
troubled	0,27	hideous	0,28
complain	0,24	lousy	0,28
flat	0,19	wine	0,26
mistake	0,18	disappointing	0,23
problems	0,18	vegetarian	0,20
ankle	0,17	tasted	0,17

Table 5 shows the association between words in negative reviews, they are the topics most frequently discussed by visitors in their review. Based on these tables can be obtained some of the following information.

Words associated with the word "service" in negative reviews provide information about customer or passenger complaints about poor service, silly, twisted, sad, confusing, failure, schedule delays, pricing issues and compensation.

The words associated with the word "staff" in the negative review provide information about customer or passenger complaints about the performance of Garuda Indonesia airline staff who are considered confused, unfriendly, chaotic, unanswered and lack of communication and information delivery resulting in misunderstandings.

The words associated with the word "seat" in negative reviews provide information about the full, narrower, smaller, worse seats, ankle spots and many who complain of the separated distance between their companions.

The words associated with the word "food" in the negative review provide information about some of the passengers' satisfaction with the food served by Garuda Indonesia maskpai which is considered disappointing, appalling, taste problem, portion, so most are not consumed. In addition passengers complained about the food provided mostly local menus and no wine or wine.

CONCLUSION

Based on the results of the analysis that has been done, obtained some conclusions as follows:

1. By using the comparison of trainer data and test data of 80%: 20% obtained the classification of sentiments using the Naïve Bayes Classifier model obtained an accuracy of 82.02%, which means that of 228 tested review data, there are 187 reviews are correct classification.
2. Based on the results of classification and text associations conducted, it is generally known that Garuda Indonesia airline passengers mostly discuss about service, staff, and food because it always appears in both positive and negative sentiment class. In general, the text association method used shows the extraction of information in positive classes such as service, food, seat, time, staff, entertainment, check-in, and cabin. While in the negative class that is often complained of including service, staff, seats, food, hour, check-in, luggage, and boarding.

ACKNOWLEDGEMENTS

Academic Development Directorate of Islamic University of Indonesia (BPA UII) and Directorate of Talents, Interests, and Student Welfare for funding the conference.

REFERENCES

Baeza-Yates, R., and B. Ribeiro-Neto, "Modern Information Retrieval", New York: ACM Press, 1999.

Dave, K., Lawrence, S., dan Pennock, D.M., "Mining the peanut gallery: opinion extraction and semantic classification of product reviews", Hungary: ACM, Proceeding of WWW '03 Proceedings of the 12th international conference on World Wide Web (pp. 519-528), 2003.

Hu and Liu., 2004, "Opinion Lexicon: A list of English positive and negative opinion words or sentiment words", http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar.

Lee, L., dan Pang, B., "Opinion Mining and Sentiment Analysis", Foundation and Trends in Information Retrieval, 2(1-2): 1-135, 2008.

Manning, C. D., et.al., "An Introduction to Information Retrieval – Online Edition", Cambridge: Cambridge University Press. 2009.

Nasukawa, T. & Yi, J., "Sentiment Analysis: Capturing Favorability Using Natural Language Processing". In Proceedings of the 2nd International Conference on Knowledge Capture. pp. 70–77., 2003.

Suthaharan, Shan., "Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning", Springer. p. 10. ISBN 9781489976413, 2015.

Tan, et al., "Introduction To Data Mining". USA: Addison-Wesley, 2006.

Tjiptono, F., "Service Quality Satisfaction". Yogyakarta: Andi. 2005.